<u>Remarks</u>

In the Office Action, the Examiner rejects claims 1-31 under 35 U.S.C. § 103(a) as being unpatentable over the PCT International Publication Number WO 03/017023 to <u>Galai et al.</u> ("Galai") in view of U.S. Patent Number 6,665,658 to <u>DeCosta et al.</u> ("DeCosta").

By this Amendment, Applicants amend claims 1, 10, 21, and 26 and cancels claims 9, 11, 22, 27, and 31 without prejudice or disclaimer. Claims 1, 10, 21, and 26 are amended to include certain features from canceled claims 9, 11, 22, and 27, respectfully.

Claims 1-8, 10, 12-21, 23-26, and 28-30 remain pending.

Applicants note that an interview was conducted with Examiner Rutledge on October 3, 2006. Applicants appreciate the courtesy extended by the Examiner during the interview. In the interview, Applicants' representative particularly discussed Galai and the differences between Galai and the present invention. Additionally, Applicants proposed certain claim amendments to further clarify the differences between Galai and the invention. The Examiner indicated that she would like to further review Galai before making a final determination on the allowability of the claims.

Claims 1-31 stand rejected under 35 U.S.C. § 103(a) based on Galai and DaCosta. Initially, Applicants note that the rejection of claims 9, 11, 22, 27 and 31 are obviated by virtue of their cancellation. For the following reasons, Applicants respectfully disagree with the rejections of pending claims 1-8, 10, 12-21, 23-26, and 28-30.

Claim 1, as amended, is directed to a method for crawling documents comprising receiving a uniform resource locator (URL); receiving at least two different copies of a document associated with the URL; and determining whether a web site corresponding to the URL uses session identifiers based on a comparison of URLs that are within the document and that change between the at least two different copies of the document, where the web site is determined to use session identifiers when a portion of the URLs that change between the at least two different copies of the document is greater than a threshold.

In rejecting claim 1, the Examiner contends that Galai discloses many of the features recited in claim 1, but concedes that Galai does not disclose "that the URLs are compared for the specific purpose of determining whether the web site uses session identifiers." (Office Action, sentence bridging pages 3 and 4). For this, the Examiner relies on DaCosta. (Office Action, page 4).

Applicants respectfully disagree with the Examiner's interpretation of Galai with regard to amended claim 1. As noted by the Examiner, Galai appears to disclose comparing a web page with a second web page, which was accessed with a reduced version of the URL used to access the first web page, to determine if the two web pages are similar. (Office Action, page 3 and Galai, page 20, lines 13-20). Galai notes that if the two web pages are similar, this may indicate that the parameter used to reduce the URL is redundant. (Galai, page 20, lines 21 and 22). Galai clearly discloses determining whether the parameter used to reduce the URL is redundant based on a comparison of the two web pages. Galai discloses additional details about the techniques used for comparing web pages at page 21. (See Galai, page 21, lines 5-20). For

10

example, Galai discloses that the web page comparison function may be based on a comparison for similarity in content or a comparison for visual similarity.

Claim 1, in contrast to Galai, recites, among other things, determining whether a web site corresponding to a URL uses session identifiers "based on a comparison of URLs that are within the document and that change between the at least two different copies of the document, where the web site is determined to use session identifiers when a portion of the URLs that change between the at least two different copies of the document is greater than a threshold." Galai does not disclose or suggest comparing URLs that are within a document in the manner recited in claim 1. In particular, Galai does not disclose or suggest a comparison in which "the web site is determined to use session identifiers when a portion of the URLs that change between the at least two different copies of the document is greater than a threshold," as recited in claim 1. (emphasis added). In contrast, Galai explicitly discloses comparing two web pages for similarity in content or in visual similarity. The Examiner can appreciate that comparing web pages for similarity in content or visual similarity is significantly different than determining when a portion of the URLs that change between the at least two different copies of the document is greater than a threshold, as recited in amended claim 1. Galai does not specifically disclose comparing URLs, much less comparing URLs to make the determination recited in claim 1.

Claim 9, which is now canceled, previously recited features similar to those currently recited in claim 1. In rejecting claim 9, the Examiner points to page 27, line 6 through page 28, line 21 of Galai as allegedly disclosing the features of the previous version of claim 9. (Office Action, paragraph bridging

11

pages 5 and 6). These sections of Galai disclose material similar to that

discussed previously at pages 21 and 22 of Galai. More specifically, these

sections of Galai disclose:

> As shown, in stage 1, the Web page is preferably retrieved by using
> the complete URL to form an original Web page. In stage 2, each of
> the parameters is preferably removed and the Web page is
> retrieved again by using the reduced URL. The term "parameter"
> refers to any divisible subunit of the URL. In stage 3, this Web page
> is then compared with the original Web page. If the removed
> parameter (s) are not redundant, such that they are required for the
> correct retrieval of the original Web page, then the retrieved Web
> page would be completely different from the original Web page.

> If the parameter is redundant, the Web pages may be expected to
> be similar, although perhaps not completely identical. Lack of
> identity may occur if the Web page includes one or more links with
> the complete URL, as for a session ID. Alternatively, the Web page
> could be custom tailored according to user identifying information,
> for personalization. For that reason, the comparison function of the
> present invention preferably checks for similarity in content and
> more preferably produces a similarity level, which is the likelihood
> of the two Web pages to have the same content. If this value
> exceeds a certain threshold, then most preferably the removed
> parameter is considered to be redundant.

> According to preferred embodiments of the present invention, the
> level of similarity is determined according to visual similarity. Visual
> similarity is preferably determined according to two different types
> of parameters. A first type of parameter is based upon content of
> the document, such as text and/or images for example. A second
> type of parameter is based upon visual layout characteristics of the
> document, such as the presence of one or more GUI (graphical
> user interface) gadgets or the location of text and/or images, for
> example. More preferably, the level of similarity is determined by
> comparing content-based parameters between documents, rather
> than by comparing visual layout characteristics. The use of content-
> based parameters is preferred because similarity is preferably
> determined according to the actual content or "meaning" of a
> document, with regard to being submitted to a search engine and/or
> otherwise stored. The above process preferably produces
> instructions on a process for detecting redundant parameters in
> URLs with the same structure, in order to remove these redundant
> parameters as the normalization instructions. The above process is
> preferably executed once per URL structure, and the normalization
> instructions are then applied to each URL with the same structure.

> The term "URL structure" preferably includes any part of a URL
> having the same parameters, repeated for each such structure.
> Therefore, stages 1- 3 are optionally and preferably repeated for
> each URL structure. Once a parameter and/or a URL structure has
> been identified as occurring repeatedly, optionally and preferably,
> stages 1-3 are not performed again for such repeated parameters
> and/or URL structures.

(Galai, page 27, line 6 through page 28, line 21) (emphasis added). These
sections of Galai describe comparing web pages to determine similarity in
content or visual similarity. As mentioned above, comparing web pages for
similarity in content or visual similarity is significantly different than determining
when a portion of the URLs that change between the at least two different copies
of the document is greater than a threshold, as recited in amended claim 1.

The last five sentences of the above-quoted section of Galai was
particularly pointed to by the Examiner in the rejection of claim 9. (Office Action,
page 5). This section of Galai appears to generally summarize the process of
Galai for detecting redundant parameters in URLs. Galai discloses detecting
redundant parameters in URLs. The redundant parameters detected by Galai,
however, are detected by comparing web pages for similarity in content or visual
similarity. This does not disclose or suggest the features recited in claim 1.

Applicants submit that DaCosta does not cure the above-noted
deficiencies of Galai. Accordingly, Galai and DaCosta, even if combined as the
Examiner suggests, still would not disclose or suggest each of the features
recited in amended claim 1. Accordingly, the rejection of claim 1 under § 35
U.S.C. 103(a) based on Galai and DaCosta is improper and should be
withdrawn.

The rejection of dependent claims 2-8 based on Galai and DaCosta should also be withdrawn, at least by virtue of the dependency of these claims from claim 1. These claims also recite features of their own that are not disclosed or suggested by Galai or DaCosta, either alone or in combination.

Claim 4, for example, recites that the compared URLs that change include URLs that are local to the web site. Neither Galai nor DaCosta disclose this feature of claim 4. In rejecting this claim, the Examiner appears to rely on page 4, lines 15-20 as disclosing that the method of Galai "can be applied to any web page within a site." (Office Action, page 5). Page 4, lines 15-20 of Galai merely discloses a list of document types for which a search engine may process URIs. This section of Galai, however, cannot be said to disclose or suggest that the compared URIs are URIs that are local to a web site. For at least this reason also, the rejection of claim 4 is improper and should be withdrawn.

Independent claim 10 and its dependent claims 12-14 also stand rejected under 35 U.S.C. § 103(a) based on Galai and DaCosta. Applicants respectfully traverse this rejection.

Claim 10 is directed to a method for identifying web sites that use session identifiers. The method includes downloading at least two different copies of at least one document from a web site; extracting uniform resource locators (URLs) from the two different copies of the web document; comparing the extracted URLs of the two different copies of the document; and determining whether the web site uses session identifiers when the comparison indicates that at least a portion of the URLs change between the two different copies.

In rejecting claim 10, the Examiner uses rationale similar to that given when rejecting claim 1. Specifically, the Examiner contends that Galai discloses many of the features recited in claim 10, but concedes that Galai does not disclose "that the URLs are compared for the specific purpose of determining whether the web site uses session identifiers." (Office Action, page 6). For this, the Examiner relies on DaCosta. (Office Action, page 7).

Applicants respectfully disagree with the Examiner's interpretation of Galai. Galai does not disclose or suggest, as is recited in amended claim 10, extracting URLs from the two different copies of a web document, comparing the extracted URLs of the two different copies of the document, and determining whether the web site uses session identifiers when the comparison indicates that at least a portion of the URLs change between the two different copies. As previously discussed, Galai notes that if the two web pages are similar, this may indicate that a parameter used to reduce the URL through which the second web page was obtained is redundant. (Galai, page 20, lines 21 and 22). Galai clearly discloses determining whether the parameter used to reduce the URL is redundant based on a comparison of the two web pages. Comparing web pages for similarity in content or visual similarity, as described by Galai, does not disclose or suggest the features of claim 10, which include comparing the extracted URLs of the two different copies of the document and determining whether the web site uses session identifiers when the comparison indicates that at least a portion of the URLs change between the two different copies.

Applicants submit that DaCosta does not cure the above-noted deficiencies of Galai. Accordingly, Galai and DaCosta, even if combined as the

Examiner suggests, still would not disclose or suggest each of the features recited in amended claim 10. Accordingly, the rejection of claim 10 under § 35 U.S.C. 103(a) based on Galai and DaCosta is improper and should be withdrawn.

The rejection of dependent claims 12-14 based on Galai and DaCosta should also be withdrawn, at least by virtue of the dependency of these claims from claim 10. These claims also recite features of their own that are not disclosed or suggested by Galai or DaCosta, either alone or in combination.

Claim 12, for example, recites that extracting URLs from the two different copies of the document includes extracting only URLs that are local to the web site. For reasons similar to that given above with respect to claim 4, Applicants submit that neither Galai nor DaCosta disclose this feature of claim 12.

Independent claim 15 and its dependent claims 16-20 also stand rejected under 35 U.S.C. § 103(a) based on Galai and DaCosta. Applicants respectfully traverse this rejection.

Independent claim 15 is directed to a device comprising a spider component and a session identifier component. The spider component is configured to crawl web documents associated with at least one web site. The session identifier component is configured to determine whether the web site uses session identifiers based on a comparison of a portion of the URLs that change between different copies of at least one web document downloaded from the web site.

In rejecting claim 15, the Examiner uses rationale similar to that given when rejecting claims 1 and 10. Specifically, the Examiner contends that Galai

16

discloses many of the features recited in claim 15, but concedes that Galai does not disclose "that the URLs are compared for the specific purpose of determining whether the web site uses session identifiers." (Office Action, page 9). For this, the Examiner relies on DaCosta. (Office Action, page 9).

Applicants respectfully disagree with the Examiner's interpretation of Galai. Galai does not disclose or suggest, as is recited in claim 15, a session identifier component configured to determine whether the web site uses session identifiers based on a comparison of a portion of the URLs that change between different copies of at least one web document downloaded from the web site. As previously discussed, Galai discloses determining whether a parameter used to reduce the URL is redundant based on a comparison of the two web pages. Comparing web pages for similarity in content or visual similarity, as described by Galai, does not disclose or suggest the session identifier component recited in claim 15, which compares a portion of URLs that change between different copies of at least one web document downloaded from the web site.

Applicants submit that DaCosta does not cure the above-noted deficiencies of Galai. Accordingly, Galai and DaCosta, even if combined as the Examiner suggests, still would not disclose or suggest each of the features recited in amended claim 15. Accordingly, the rejection of claim 15 under § 35 U.S.C. 103(a) based on Galai and DaCosta is improper and should be withdrawn.

The rejection of dependent claims 16-20 based on Galai and DaCosta should also be withdrawn, at least by virtue of the dependency of these claims

from claim 15. These claims also recite features of their own that are not disclosed or suggested by Galai or DaCosta, either alone or in combination.

Claim 19, for example, recites that the portion of the URLs that change are identified from URLs that are local to the web site. For reasons similar to that given above with respect to claims 4 and 10, Applicants submit that neither Galai nor DaCosta disclose this feature of claim 19.

Amended independent claim 21 and its dependent claims 23-25 also stand rejected under 35 U.S.C. § 103(a) based on Galai and DaCosta. Claim 21 recites features similar to, although of different scope than, those recited in claim 10. Accordingly, based on rationale similar to that presented with regard to claim 10, Applicants submit that the rejection of claims 21 and 23-25 based on Galai and DaCosta is improper and should be withdrawn.

Amended independent claim 26 and its dependent claims 28-30 also stand rejected under 35 U.S.C. § 103(a) based on Galai and DaCosta. Claim 26 recites features similar to, although of different scope than, those recited in claim 10. Accordingly, based on rationale similar to that presented with regard to claim 10, Applicants submit that the rejection of claims 26 and 28-30 based on Galai and DaCosta is improper and should be withdrawn.

In view of the foregoing amendments and remarks, Applicants respectfully request the Examiner's reconsideration of this application, and the timely allowance of the pending claims.

To the extent necessary, a petition for an extension of time under 37 CFR 1.136 is hereby made. Please charge any shortage in fees due in connection

with the filing of this paper, including extension of time fees, to Deposit Account

No. 50-1070 and please credit any excess fees to such deposit account.

Respectfully submitted,

HARRITY SNYDER, L.L.P.

By: /Brian E. Ledell/
     Brian E. Ledell
     Reg. No. 42,784

11350 Random Hills Road
Suite 600
Fairfax, Virginia 22030
(571) 432-0800

Date: October 27, 2006